Preprint
UCRL-JC-147778

The Application of Range Space Operations to Color Images

C. Baldwin, M. Duchaineau

This article was submitted to
The Institute of Electrical and Electronics Engineers, Inc.
Visualization 2002, Boston, MA, October 27, November 1, 2002

March 26, 2002

U.S. Department of Energy

Lawrence
Livermore
National
Laboratory

The Application of Range Space Operations to Color Images

Chuck Baldwin Mark Duchaineau

The Center for Applied Scientific Computing
Lawrence Livermore National Laboratory

Abstract

The knowledge gained from scientific observation, experiment, and simulation is linked to the ability to analyze, understand, and manage the generated results. These abilities are increasingly at odds with the current, and future, capabilities to generate enormous quantities of raw scientific and engineering data from instruments, sensors, and computers. Many researchers are currently engaged in activities that seek to create new and novel methods for analyzing, understanding, and managing these vast collections of data. In this work, we present some of our research in addressing a particular type of problem in this broad undertaking. Much the scientific data of interest is in the form of observed, measured, or computed multivariate or multi-component vector field data – with either as physical or color data values. We are currently researching methods and techniques for working with this type of vector data through the use of a novel analysis technique. Our basic approach is to work with the vector field data in its natural physical or color space. When the data is viewed as a functional mapping of a domain, usually an index space, to a range, the physical or color values, potentially interesting characteristics of the data present themselves. These characteristics are useful in analyzing the vector fields based on quantities and qualities of the physical or color data values themselves. We will present the basic development of the idea of *range space* operations and detail the information we are interested in and some of the issues involved in its computation. The data we are first interested in, and discuss exclusively in this work, is color image data from scientific observations and simulations. Some of the operations on the range space representation that are of interest to this color image data are colormap construction, segmentation, color modeling, and compression. We will show some how some of the operations can be implemented in range space, what analysis capabilities they provide, and how they work on some example images. We will also discuss some of the future goals of the research along with what has been learned from this work.

CR Categories: E.4 [Coding and Information Theory]: Data compaction and compression—; I.1.5 [Pattern Recognition]: Models—Structural; I.3.m [Computer Graphics]: Miscellaneous—;

Keywords: Color Image Analysis, Modeling

1 Introduction

The problem we are interested in addressing involves a variety of analysis operations on data obtained from scientific experiment, simulation, and observation. The results from these scientific endeavors is data that is usually very large and complex, in the form of multi-variate and multi-component vector field quantities. A vector field data value can be viewed as a function $\vec{f}$ which maps a given domain $\mathcal{D}$ onto some arbitrary range $\mathcal{R}$, i.e. :

$$\vec{f}(\mathcal{D}) \mapsto \mathcal{R}. \quad (1)$$

For instance, color image data can be viewed as a function which maps two-dimensional pixel data, given as integer indexes, to the three-dimensional color values, given as unsigned integers. The general research approach we are pursuing is to analyze the data directly in this *range space*. There are several reasons for analyzing such data in its range space. The first, probably most important, reason is to find and exploit patterns and correlations in the range space — where a scientist or engineer naturally "thinks" about the data. The second, closely related, reason is to alleviate any induced patterns and correlations resulting from analysis operations in the domain space. A third reason is to take advantage of replications in the individual data elements to obtain a reduction in size (known as *dimensionality reduction*) – this is most applicable to color images. To be objective, there are issues with this approach. The first issue is that the range space data always forms an irregular point set and this must be organized and stored. Secondly, there is not extensive theory or numerous algorithms for analyzing this irregular range space data. Finally, even though there may be a reduction in the *data dimensionality*there may be an increase in the *physical dimensionality* of the individual data elements – again this is most applicable to color images. Further, the savings are problematic if some domain space information needs to be preserved.

Given these arguments, we have decided to explore various applications of this range space mapping where it seams applicable. That is, where some general analysis of the data values might add significantly to the understanding of some phenomena. The most obvious applications occur in color image data from scientific experiments, simulations and observations. Here we are interested in several range space operations applied to the image data :

- creating simplified color models of the image data,
- obtaining clusters of the image data in color space,
- computing segmentations of the image data based on color values,
- constructing a color table,
- general compression of the image data.

Creating color models and clusters of image data in range space are of interest for reducing the complexity of the data – ideally to a few simple linear functions or relationships. Computing segmentations based on color are of interest because the color data itself is often directly related to a field quantity. Constructing a color table is often a simple matter once the range space data is computed and stored and is of interest as it serves as a useful global statistical analysis tool. Compression is always of interest in large scale scientific computing applications and, as with segmentation, the color data is the information about a field quantity. It should be noted that the ideas presented for range space operations on color image data can be extended to other types of data, such as vector velocity fields.

To further this general methodology we are researching both the theory and algorithms needed for analyzing scientific image data based on range space behavior. We will detail some of the results obtained thus far for this endeavor. In Section 2 we give the general concepts and algorithms for computing and storing the range space representation of color images. In Section 3 we will detail the ideas behind using the representation to compute two operations mentioned above, construction of a color table and color segmentation. Finally, in Section 4 we will discuss what we have learned so far and how the ideas can be expanded upon to solve some of the more complex operations, such as modeling and compression, we are trying to create.

2 Algorithms

The first task is to compute the range space of a given image and establish some notions about the quantities that make up the stored range space. We begin with a color image which we will indicate functionally as $\vec{I}(i,j)$. Note that we will be using red, green, blue (RGB) color values but the analysis follows for any color space. The domain of $\vec{I}(i,j)$ is a rectilinear *mesh* given by :

$$(i,j) : 1 \leq i \leq n_i, 1 \leq j \leq n_j, \tag{2}$$

where n_i and n_j are the respective sizes of the mesh in the i and j directions. The range of $\vec{I}(i,j)$ is a collection of red, green, and blue color values given by :

$$\vec{I}(i,j) = \begin{bmatrix} r(i,j) \\ g(i,j) \\ b(i,j) \end{bmatrix} : \begin{matrix} 0 \leq r(i,j) \leq s_r \\ 0 \leq g(i,j) \leq s_g \\ 0 \leq b(i,j) \leq s_b \end{matrix}, \tag{3}$$

where s_r, s_g, and s_b are the maximum size of red, green, and blue color values. To compute the range space of $\vec{I}(i,j)$ we effectively need to construct its colormap. In constructing the point set associated with the range space of the image we, in addition to the color value, wish to keep the number of times each color value occurs in the image. This will prove useful in many operations involving the range space. The information we keep for each unique value in range space is given in table 1.

Variable	Name
ρ	red color value
γ	green color value
β	blue color value
ω	frequency

Figure 1: Information associated with a range space point

The range space value :

$$\begin{pmatrix} \rho \\ \gamma \\ \beta \end{pmatrix}, \tag{4}$$

is a unique color value for the image while ω is its frequency in the image. The basic algorithm to compute the collection of color value points, the range space *point set*, is :

- initialize a data store for the range space data
- for each pixel in the original image
  - search for the color value in the range space data
  - if it exists, increment its frequency; otherwise add it

Currently, we store the information in table 1 in a sorted list, ordered by decreasing frequency, which defines an irregular point set in the range space. This effectively yields a set of range space points $\mathcal{R}$ given by :

$$\mathcal{R} = \left\{ \begin{pmatrix} \rho_k \\ \gamma_k \\ \beta_k \\ \omega_k \end{pmatrix} \right\}_{k=1}^{N}, \tag{5}$$

where N is the number of unique range space points. There are some efficiency issues with this list based algorithm, associated with the searching portion of the basic algorithm. However, some operations on this set we are currently exploring use this fundamental structure (augmented to become a graph vertex) to establish graphs among collections of the range space points. We are also interested in exploring a new algorithm for more efficiently computing the range space points $\mathcal{R}$. This two step process is :

1. Initially store the range space points in a hash table using the (ρ, γ, β) value as a hash key,
2. Extract the range space points and sort them in ascending order based on frequency.

Since we would only doing dictionary like operations on the hash table, *insert* and *search*, an open hashing technique [1] should prove an effective solution to the construction efficiencies, once a suitable hashing function is defined. We present three examples of the results of this range space computation procedure. The first pair of figures 2 shows an image of a spiral galaxy and its point set in range space. The second pair of figures 3 shows an image from a the visualization of a hydrodynamics simulation and its point set in range space. The third pair of figures 4 shows an image from the visualization of a laser simulation and its point set in range space.

3 Applications

With the point set of the image is computed in range space we can either directly use the information or augment it to obtain several operations of interest. The computation of a color map for an image is one operation that is easily performed with the basic range space data. The colormap is a structure that simply indexes the unique colors of an image – the data in range space. This can be stored in table form after the color value point set is computed. If an index image is to be computed, whereby the values of the individual image pixels are indexes into the colormap, the basic algorithm presented earlier can be augmented to carry the pointer (or index) into the range space point set data.

An interesting use for the range space data is to obtain simplified models of the color space data and use that for basic analysis, segmentation, or compression operations. The problem is modeling collections of range space values is generally a difficult and time consuming process. For instance, looking at figures 2, 3, and 4 it is easy to see visual patterns and possible simplifying models but computing them is considerably harder. We are researching different techniques for computing these simplified models of global range space behavior. The method we are currently focusing on is a geometric method based on neighborhood operations [3, 4] in the original image domain and color-space distances of the neighboring pixel values. The goal in this technique is to establish a simple model of global behavior based on locality in image space and proximity of color in color space. This technique starts with a standard 8-point neighborhood of a pixel $p(i,j)$ at location (i,j) in the images domain given by :

$$\left\{ \begin{array}{ccc} p(i-1,j-1) & p(i,j-1) & p(i+1,j-1) \\ p(i-1,j) & & p(i+1,j) \\ p(i-1,j+1) & p(i,j+1) & p(i+1,j+1) \end{array} \right\} \quad (6)$$

For each neighboring pixel of $p(i,j)$ in image space we compute the distance of the corresponding color space values with an applicable norm and compare against a user supplied tolerance ϵ. If the distance is smaller than ϵ a segment is attached between the range space values. This creates a complex undirected graph in range space for the image. An example of this graph for the image in figure 2 is given in figure 5. The resulting graph is too complex to use as a model but the geometric structure can be simplified to linear structures (points, lines, and planes) by techniques similar to those used in the work of Garland and Heckbert [2] and Lindstrom [5]. After simplification we obtain two linear models for the range space data given in figures 6 and 7. Using these two models as a basis for two new colormaps, and thus images, we obtain an interesting segmentation of the original image based on range space characteristics. Figures 8 and 9 are the results of the range space segmentation based on the two models from the image in figure 2. The segmentation shows that the first model is composed of pixels in the bright core of the galaxy in the original image and the second is composed of the pixels in the galaxies sprial arms in the original image. This segmentation is of interest to astronomers examining this data, as the color values usually have a physical interpretation to the scientist.

4 Future Research

We have shown how the basic range space idea can computed and applied to two interesting problems of scientific data processing. There are many avenues of research which we are currently investigating in detail or have plans to explore further. Currently under investigation are the new, and faster method for computing the range space of an image. As outlined in Section 2, this method uses hash functions to compute the basic range space data structures needed. This has implications for the efficiency of other operations using the structure though. Our main research objective is in modeling the range space data in an effective and efficient way. We presented one algorithm based on geometric (neighborhood) considerations for computing models. Interesting methods based on clustering or general least squares techniques applied to the range space data are future areas of potentially useful exploration. The clustering algorithms are particularly applicable, since many are are designed with this irregular point set case in mind. Using the modeling as a basis for compression is a avenue of research we are also actively exploring. The idea is to use the models as reduced set representations for the image, perhaps using a few representatives from the model as estimates for large collections of "like" pixels. In the application of these techniques, we have one particular goal in mind. The development here has been with respect to color images, however any vector field can potentially benefit from this range space mapping. One of these vector fields which we deal with on a regular basis is velocity fields from scientific simulations. Applying, with adaptations, the ideas presented here can be of considerable advantage to scientists in analyzing velocity field behavior.

5 Conclusions

In this paper we have presented ongoing research into new and novel techniques for analyzing vector field data. We have established a methods for computing the range space of a vector field by treating the data as a functional mapping and examining the resulting map by its range behavior. Using this methodology, we have created a representation of the resulting irregular point set and pursued various operations on the representation. We presented how the operation of colormap computation, or retrieval, could easily be implemented using the range space representation. We have also presented results on how modeling techniques could be used on the range space representation to define simplified models of the complex range space behavior. The effect of this modeling has been shown through the use of a color segmentation of a target image. We feel that the techniques and methods presented shows that the idea of range space representations and operations are a beneficial tool for scientists and engineers as well as general researchers using image data.

References

[1] Thomas H. Cormen, Charles E. Leiserson, and Ronald L. Rivest. *Introduction to Algorithms*. McGraw-Hill, 1989.

[2] Michael Garland and Paul S. Heckbert. Surface simplification using quadric error metrics. In *Proceedings of SIGGRAPH 97*, pages 209–216, 1997.

[3] Jähne, B. *Practical Handbook on Image Processing for Scientific Applications*. CRC Press, Boca Raton, Florida, 1997.

[4] A. K. Jain. *Fundamentals of Digital Image Processing*. Prentice Hall, Englewood Cliffs, New Jersey, 1989.

[5] Peter Lindstrom. Out-of-core simplification of large polygonal models. In *Proceedings of SIGGRAPH 2000*, pages 259–262, 2000.

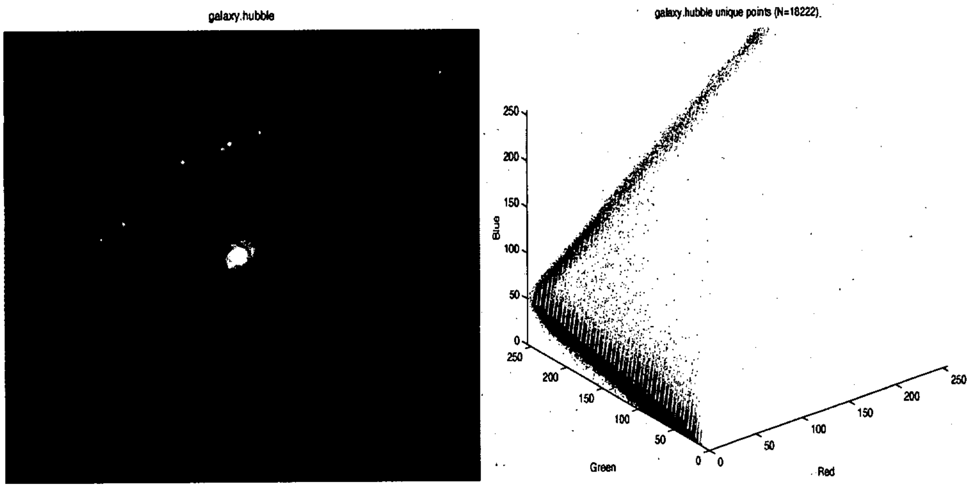

Figure 2: Original galaxy image and range space data distribution.

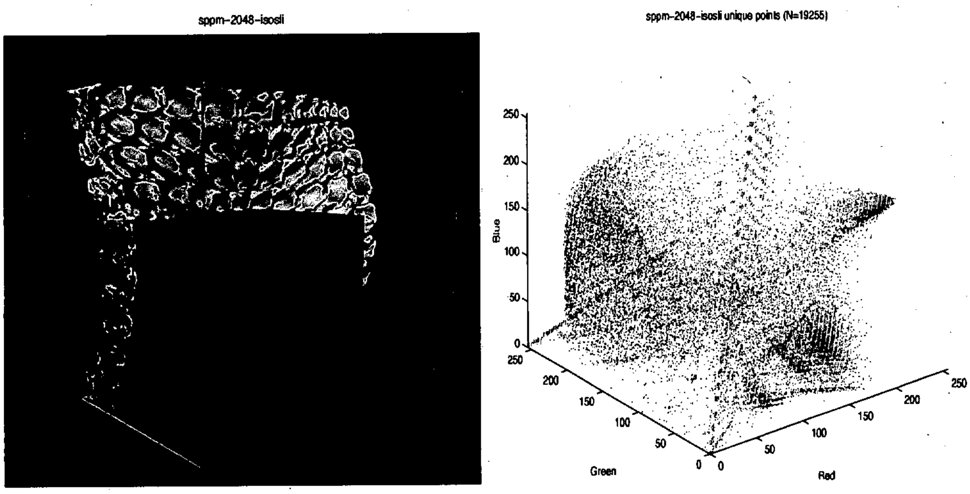

Figure 3: Original fluid mixing image and range space data distribution.

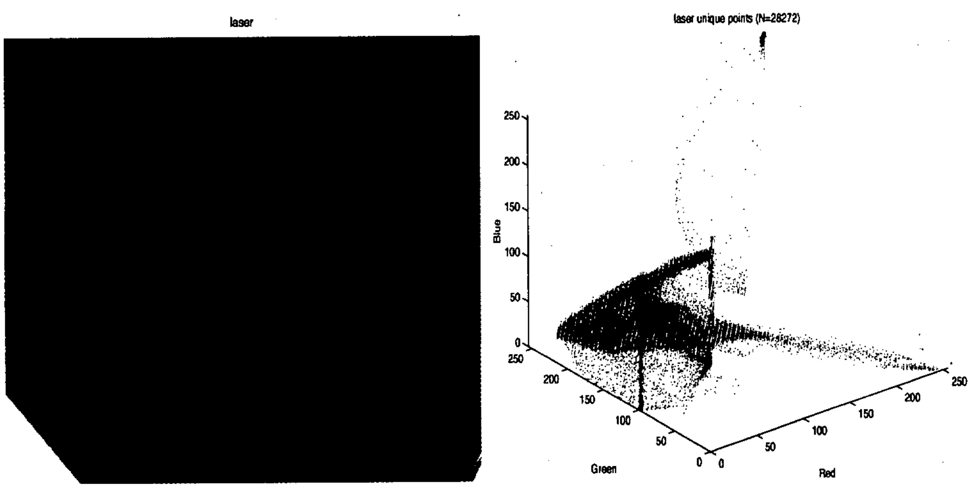

Figure 4: Original laser beam image and range space data distribution.

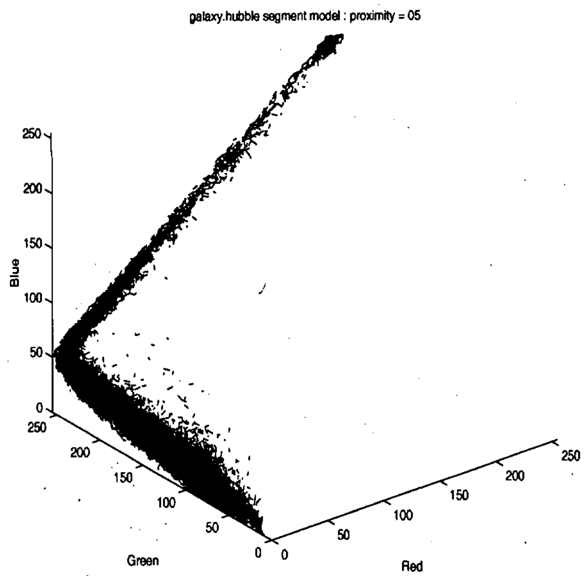

Figure 5: Segment model of galaxy image using a threshold of 5.

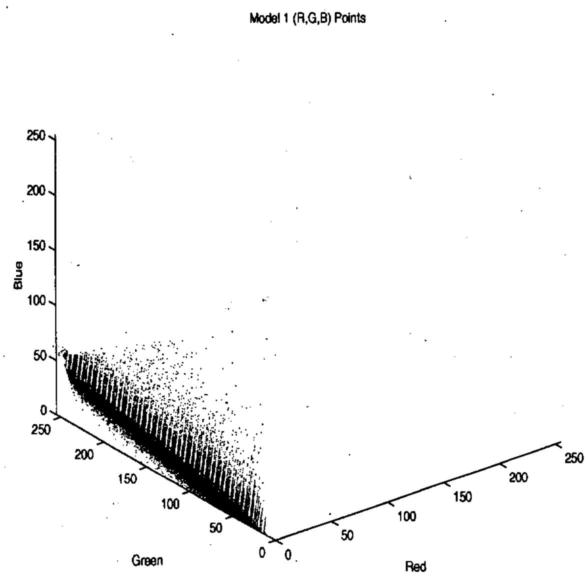

Figure 6: Range space of model 1.

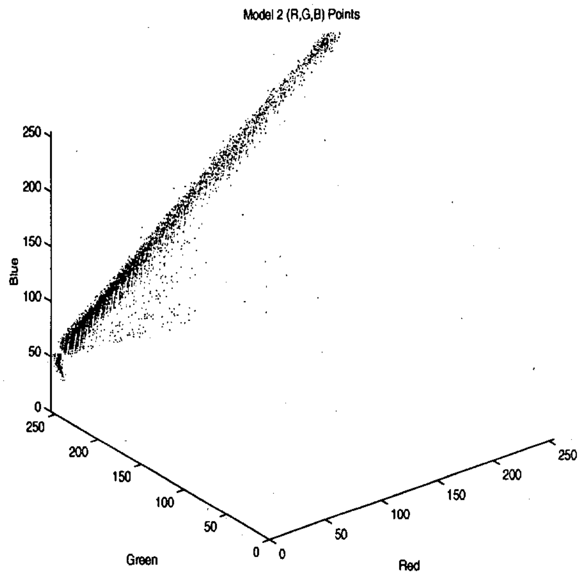

Figure 7: Range space of model 2.

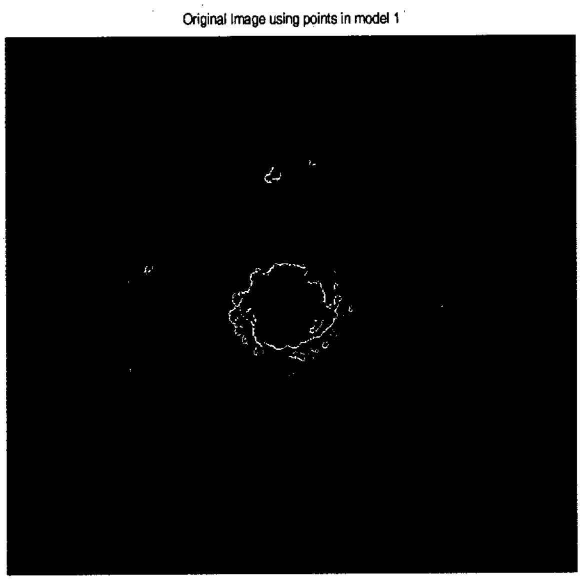

Figure 8: Galaxy image using color values from model 1.

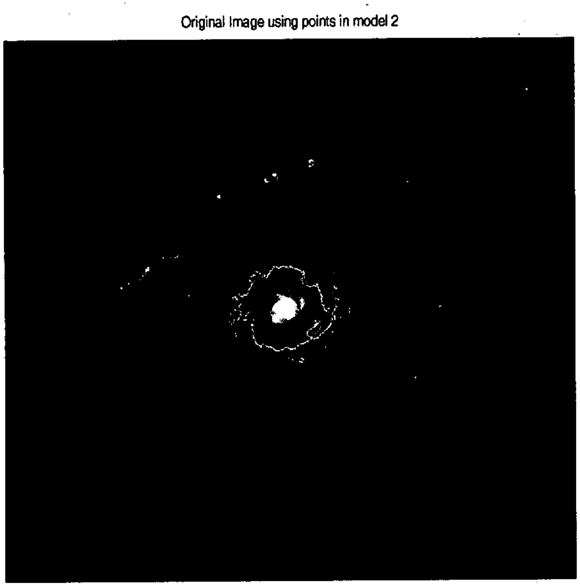

Figure 9: Galaxy image using color values from model 2.